IOWG Session

Feb 7th

Tuesday Afternoon

12:45-1:25 Rick Stevens, TeraGrid

1:25-2:15 BRC update/discussion

Break

2:30-2:45 Ross Overbeek, Related Genomes

2:45-4:00 SOP discussion

• TIGR + hegemony → 234

- TIGR + hegemony → 234
- TIGR + hubris → 316

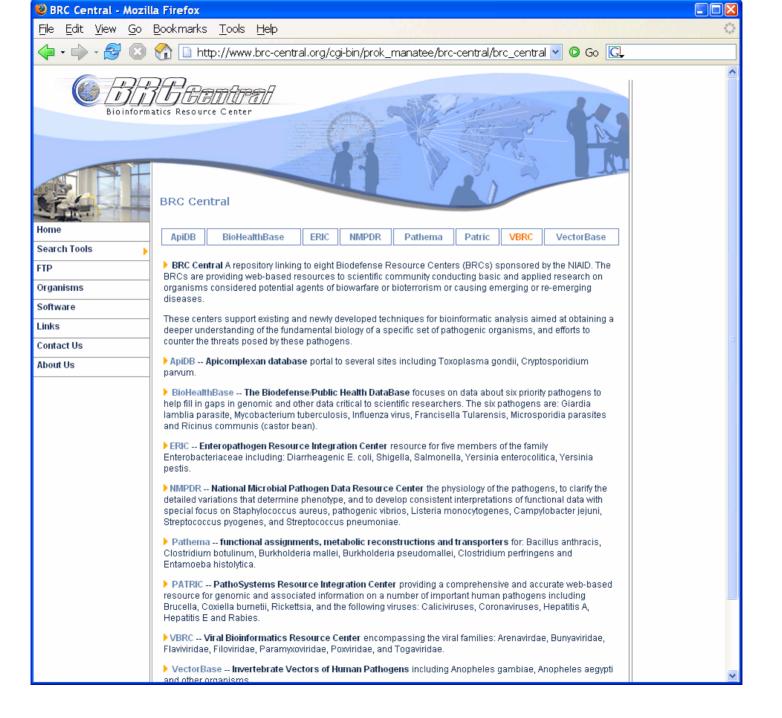
- TIGR + hegemony → 234
- TIGR + hubris → 316
- Owen + White + incompetent → 108,000

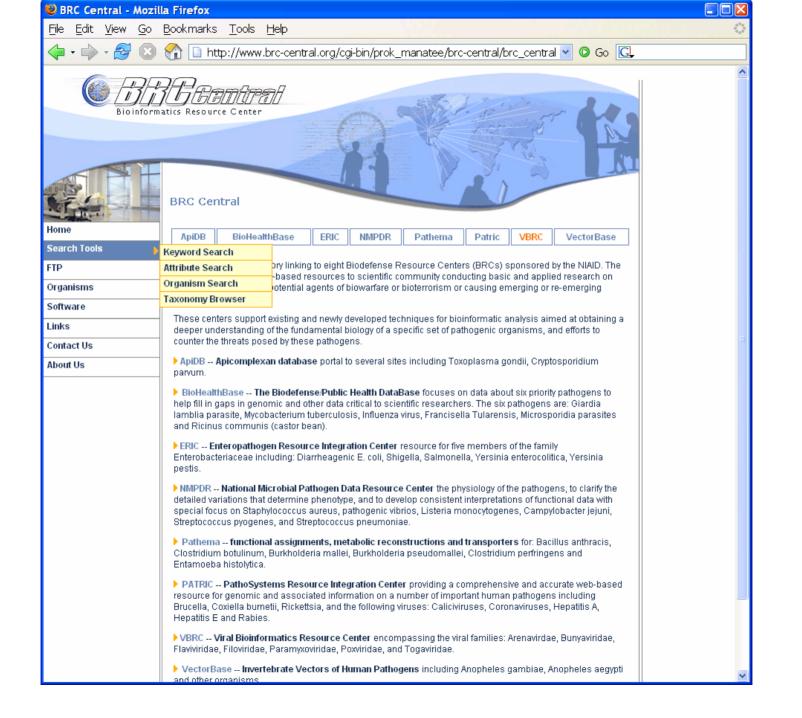
- TIGR + hegemony → 234
- TIGR + hubris → 316
- Owen + White + incompetent → 108,000
- Bruno + Sobral + incompetent → 28



BRC-central Usage

BRC Central					
Month	Unique Visitors	Number of visits	Page visits	Bandwidth	
Nov 2005	78	98	240	8.75 MB	
Dec 2005	105	147	316	13.51 MB	
Jan 2006	53	78	261	14.39 MB	

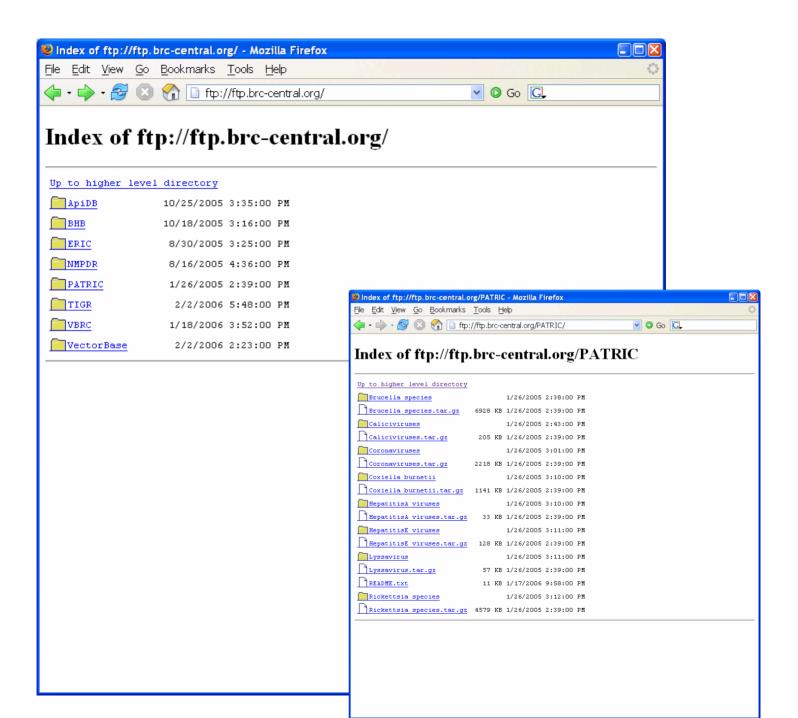




Home I	FTD I	Linke I	Contact He

Category	ApiDB	BioHealthBase	ERIC	NMPDR	
Α	-	Francisella tularensis	Yersinia pestis	-	
В	Toxoplasma gondii, Cryptosporidium parvum, Plasmodium phylum	Giardia lamblia, Microsporidia, Ricinus communis	Diarrheagenic E. coli, Yersinia enterocolitica, Shigella, Salmonella	Staphylococcus aureus, Pathogenic Vibiros, Listeria monocytogenes, Campylobacter jejuni, Streptococcus pyogenes, Streptococcus pneumoniae	
С	-	Mycobacterium tuberculosis, Influenza Virus	-	-	

Category	Pathema	Patric	VectorBase	VBRC
А	Bacillus anthracis, Clostridium botulinum		-	Variola major Virus, Arenavirus, Hanta Virus, Rift Valley Fever Virus, Ebola Virus, Marburg Virus, Dengue Virus
В	Burkholderia mallei, Burkholderia pseudomallei, Clostridium perfingens, Entamoeba histolytica	Rickettsiae, Brucella, Coxiella burnetii, Calicivirus, Hepatitis A Virus	-	California encephalitis group Virus, Kyansanar forest disease Virus, Omsk hemorrhagic fever Virus, West Nile Virus, Alphavirus
С	-	Rabies Virus, Coronavirus	Anopheles gambiae, Aedes aegypti, Anopheles gunestus, Culex pipiens, Ixodes scapularis	Hantaan Virus, Puumala Virus, Crimean-Congo hemorrhagic fever Virus, Yellow fever Virus, Tick-borne Encephalitis, Nipah Virus, Equine morbillivirus





Software page

example

Software Tags

- Software Name
- Institution
- Description
- Headline
- Creator
- Contact
- Web Site
- Open Source
- Download Site
- Current Version
- Current Version Release Date
- Relevant Publications
- Platforms
- Requirements
- Other Information
- Keywords
- Listing Last Updated

Software Tags

- Software Name
- Institution
- Description
- Headline
- Creator
- Contact
- Web Site
- Open Source
- Download Site
- Current Version
- Current Version Release Date
- Relevant Publications
- Platforms
- Requirements
- Other Information
- Keywords
- Listing Last Updated

- Proposal: adopt software submission xml
- Format
- Tags
- Submission strategy

Proposal: feature versioning

- GFF3 contains a stable_id attribute
- Format: accession.version (e.g., like Genbank)
- Applies to gene (and other genomic) features
- Currently, the stable_id attribute is optional for all features
- proposal: manatory stable_id for all gene features
- proposal: accessioning/numbering is responsibility of each BRC
- proposal: changes to underlying sequence spawn version changes

Note: A different mechanism deals with other types of changes

Proposal: data set versioning

- data sets version is created upon BRC Central deposition
- version number specified in GFF3 file
- associated with each genome
- multiple submission (eg. file/chromosome), get same version number
- allows for:
 - independent updates to the GFF3 data
 - capability to retrieve old data sets
- proposal: versioning is responsibility of each BRC
- proposal: storage of old data set is responsibility of each BRC
- proposal: automated check at BRC-central
 - ensure the version# is valid
 - doesn't conflict with a previous submission.

Proposal: tracking feature changes

- Change examples:
 - concept: "Last modified"
 - not: "Show full edit history"
 - changes to annotation
 - new annotations
 - features additions
- This is not versioning. Not all changes result in new accession version.
- For each new data set submission:
 - should come with a change log file to give best effort in detailing change
 - data stamps
 - elaboration of the "Whats New" RSS feed
- Proposal: any change triggers new version number for data set.

Expected capability for users

- "Retrieve annotations that have been modified..."
- "When was the last time Anthrx strain 12 was last submitted?"
- "I made primers for a gene 2 years ago, where is it?"

brc-central: new features

New features: Round Trip Validation

<u>Center</u>	Ontology_term	ec_number	description (gene_name)	Name (locus)	gene_symbol	organism_name	Dbxref:taxon
ApiDB	0	0	12831	13411	0	1454	1454
BHB	0	0	23003	30357	0	58	58
ERIC	0	2402	102106	104864	40410	947	947
NMPDR	0	0	63878	0	0	292	292
PATRIC	0	0	23991	0	5447	354	354
TIGR	67085	28469	149139	149139	45363	1309	1309
VBRC	0	0	5119	5119	0	414	414
VectorBase	0	0	0	0	0	6	6



GFF3

- Other issues
 - Ev codes → later
 - Other?

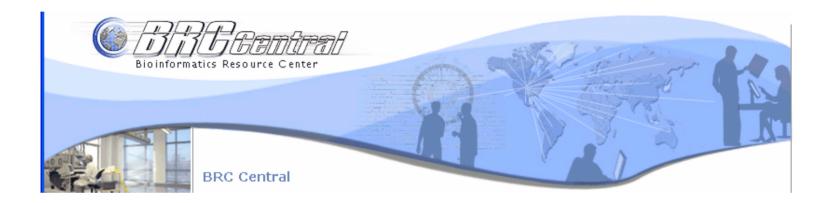
Brc-central new features

- Schedule of potential meetings
- Meetings BRCs are attending
- Round trip validation
- Blast searches
- Relationalize
- API

Decoration w/ datatypes

- pathInfo
- BHB
- OrthoMCL
- Sub-systems
- Epitopes (Immune Epitope DB and Analysis Resource)

Improve?



Suggestions for new features

- Searching
- Improved usage statistics
- Notifications
- History management

Break for Ross

Standard Operational Procedures

Feb 7th

Management by Standard Operational Procedures (SOPs)

Software, databases and interfaces perform routine processes.

Goal: create reproducible and consistent annotation.

Management by Standard Operational Procedures

Scripts/Pipelines/Automated Analysis:

- -encourages rigorous application of processes.
- -while curation of data is possible.
- -it is still not enough.
- The *reasoning* for why processes are applied is not explicit in the system.
- The specific points of evaluation need to be identified, and documented.

Benefits of SOPs

- Constructive competition
- Human resources/MGT
 - Chaos is demoralizing
 - Enforces clarification of thought
 - Pumpkinmorphogenisis
 - Essential for training
- Scale-up
 - Knowledge of required resources
 - Coordination
 - Identifies specialized working groups

Benefits of SOPs

- Publication
 - Scientifically defendable
 - Establishes standards in the community
- Retooling the flow diagram
- Meaningful tracking of output
- Essential element of project management:
 - Exit criteria
 - Who evaluates the exit criteria







SOP examples

Engine assembly

El tigre **SOPs**

SOPs are not...

- A process diagram
- A software program or pipeline
- A user manuals
- A scientific publication (e.g., "we did our assignments according to PNAS:342...")
- Controlled vocabularies
 - Note: ev_types

SOP sharing

- Sharing SOPs will help
- Caveats, output still dependent on:
 - Hangovers
 - People
 - Training/increases in expertise
 - Quality of input information
 - Ergonomics

SOP Sharing

- Converge on types
 - Nucleotide level
 - Protein level
 - Automated assignment
 - Frameshifts
 - Vaccine targets



Proposal: Decoration w/ datatypes

- pathInfo
- BHB
- OrthoMCL
- Sub-systems
- Epitopes (Immune Epitope DB and Analysis Resource)
- SOPS: Nucleotide-level, protein-level, frameshifts
 - BRC
 - Date stamp
 - SOP version
 - NOTE: ev_codes ←~→ SOPs

Proposal: Decoration w/ datatypes

- Protein-level, implications
 - I'm talking GO
 - **\$\$**
 - Where to get help
 - ev_codes ←~→ SOPs